Probabilidade e Estatística Matemática I

Parte 1: Probabilidade

Silva, J.L.P.

Junho de 2022

Apresentação da disciplina

Ementa

- **Parte I (Probabilidade)**. Profs: *José Luiz Padilha* e *Anderson Ara*
    - Probabilidade e variáveis aleatórias, funções de probabilidade, densidade probabilidade.
    - Vetores aleatórios, distribuição conjunta, marginal e condicional.
    - Esperança, variância e covariância.
    - Principais distribuições de probabilidade univariada e multivariadas.
    - Funções de variáveis aleatórios, função geradora de momentos, cumulantes e característica.
    - Desigualdades, convergência de variáveis aleatórios e teorema do limite central.
- **Parte II (Inferência)**. Profs: *Wagner Bonat* e *Paulo Justiniano*
    - Modelagem estatística e inferência.
    - Método da máxima verossimilhança e suas propriedades.
    - Inferência Bayesiana.

Procedimentos Didáticos e Avaliação

- Disciplina dividida em 4 partes: 2 partes cobrindo Probabilidade e 2 partes cobrindo Inferência.
- Serão feitas 4 avaliações, sendo uma avaliação por professor.
- Teremos listas de exercícios que servirão como preparativos para a prova.
- O formato da disciplina será híbrido, com aulas online e avaliações presenciais.

As datas das aulas de Probabilidade serão:

- **Padilha:** 08/06 e 10/06; 15/06 e 17/06; 21/06 e 24/06 [AV1].
- **Ara:** 29/06 e 01/07; 06/07 e 08/07; 13/07 e 15/07 [AV2].

Bibliografia

Wood, S. (2015). *Core Statistics.* Cambridge: Cambridge University Press.

Silvey, S. D. (1970). *Statistical Inference.* London: Chapman & Hall.

De Groot, M. H. and M. J. Schervish. (2002). *Probability and statistics.* Boston: Addison-Wesley.

Cox, D. R. and D. V. Hinkley. (1974). *Theoretical Statistics.* London: Chapman & Hall.

Mood, A. M.; Graybill, F. A. and Boes, D. C. (1974). *Introduction to the theory of Statistics.* McGraw-Hill.

Magalhães, M. N. (2015). *Probabilidade e variáveis aleatórias.* Edusp.

Tipos de probabilidade

Definição clássica (a priori)

Definição Se um experimento aleatório pode resultar em n resultados mutuamente exclusivos e igualmente prováveis e se n_A destes resultados têm um atributo A, então a *probabilidade* de A é a fração n_A/n.

Exemplo Um dado comum é lançado - há seis resultados possíveis - e um dentre os seis números pode aparecer na face superior. Se o dado é *honesto*, os seis resultados são *igualmente prováveis* (é esperado que cada face aparecerá com igual frequência relativa em lançamentos repetidos).

Assim, a probabilidade de sair um número par será $3/6$ ou $1/2$.

Algumas dificuldades:

- Como podemos definir probabilidades se o dado não for honesto?
- Qual a probabilidade de uma pessoa morrer antes dos 50 anos?

Definição frequentista (a posteriori)

Considera frequências relativas em repetições independentes como o valor da probabilidade.

Voltemos ao exemplo do dado e imagine que repetimos o experimento de lançar o dado de forma independente e registrar a face virada para cima.

```
set.seed(123)
# 100 lançamentos de um dados
prop.table(table(sample(1:6, size = 100, replace = TRUE)))
```

```
   1    2    3    4    5    6 
0.19 0.15 0.18 0.11 0.16 0.21 
```

```
# 1.000 lançamentos de um dados
prop.table(table(sample(1:6, size = 1000, replace = TRUE)))
```

```
    1     2     3     4     5     6 
0.169 0.171 0.165 0.161 0.168 0.166 
```

Definição frequentista (a posteriori)

```r
# 10.000 lançamentos de um dados
prop.table(table(sample(1:6, size = 10000, replace = TRUE)))
```

```
     1      2      3      4      5      6 
0.1712 0.1727 0.1668 0.1584 0.1652 0.1657 
```

```r
# 100.000 lançamentos de um dados
prop.table(table(sample(1:6, size = 100000, replace = TRUE)))
```

```
      1       2       3       4       5       6 
0.16607 0.16811 0.16705 0.16618 0.16787 0.16472 
```

Note como as frequências relativas se aproximam do valor real $1/6$. A probabilidade é tomada como o *limite* das frequências relativas.

Probabilidade axiomática

Teoria de conjuntos

Cada elemento em uma coleção de objetos será chamado de *ponto* ou *elemento*. A totalidade de pontos é chamada de *espaço*, *universo*, ou *conjunto universal* e será chamado de Ω. Seja ω um elemento ou ponto em Ω.

Exemplo $\Omega = R^2$, em que R^2 é a coleção de pontos ω no plano e $\omega = (x, y)$ é qualquer par de números reais x e y.

Exemplo $\Omega = \{\text{todos os alunos da UFPR}\}$.

Se ω é um ponto ou elemento do conjunto A, escrevemos $\omega \in A$; se ω não é um elemento de A, escrevemos $\omega \notin A$.

Teoria de conjuntos: Definições

Definição Subconjunto Se todo elemento de um conjunto A é também um elemento de um conjunto B, então dizemos que A é um *subconjunto* de B, e denotamos por $A \subset B$ ou $B \supset A$.

Definição Conjuntos equivalentes Dois conjuntos A e B são *equivalentes*, ou *iguais*, se $A \subset B$ e $B \subset A$, o que será indicado por $A = B$.

Definição Conjunto vazio Se A não contém pontos, será chamado de conjunto *vazio*, e denotado por ϕ.

Definição Complemento O *complemento* de um conjunto A em relação ao espaço Ω, denotado por $\bar{A}, A^c$, ou $\Omega - A$, é o conjunto de todos os pontos que estão em Ω mas não em A.

Teoria de conjuntos: Definições

Sejam A e B dois subconjuntos de Ω.

Definição União A *união* de A e B, denotada por $A \cup B$, define o conjunto que consiste de todos os pontos que estão em A ou B ou ambos.

Definição Intersecção A *intersecção* de A e B, denotada por $A \cap B$ ou AB, define o conjunto de todos dos pontos que estão simultaneamente em A e B.

Definição Diferença O *conjunto diferença*, denotado por $A - B$, é definido como o conjunto de todos os pontos em A que não estão em B.

Teoria de conjuntos: Exemplo

Exemplo Seja $\Omega = \{(x, y) : 0 \leq x \leq 1 \text{ e } 0 \leq y \leq 1\}$. Defina os seguintes conjuntos:

$$A_1 = \{(x, y) : 0 \leq x \leq 1;\ 0 \leq y \leq 1/2\},$$
$$A_2 = \{(x, y) : 0 \leq x \leq 1/2;\ 0 \leq y \leq 1\},$$
$$A_3 = \{(x, y) : 0 \leq x \leq y \leq 1\},$$
$$A_4 = \{(x, y) : 0 \leq x \leq 1/2;\ 0 \leq y \leq 1/2\}.$$

Temos:

$$A_4 \subset A_1; \quad A_4 \subset A_2; \quad A_1 \cap A_2 = A_4;$$
$$A_2 \cup A_3 = A_4 \cup A_3; \quad A_1^c = \{(x, y) : 0 \leq x \leq 1;\ 1/2 < y \leq 1\};$$
$$A_1 - A_4 = \{(x, y) : 1/2 < x \leq 1;\ 0 \leq y \leq 1/2\}$$

Teoria de conjuntos: Exemplo

Sejam Ω, A_1, A_2 e A_3 conforme mostrados nos *diagramas de Venn* a seguir.

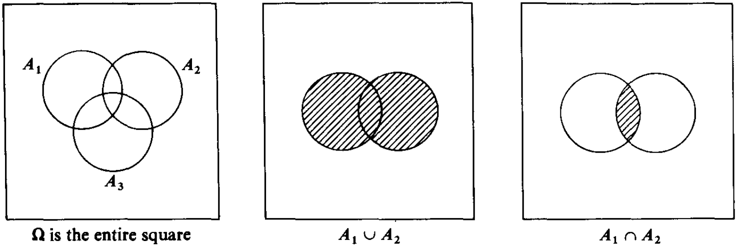
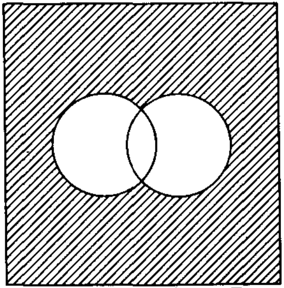
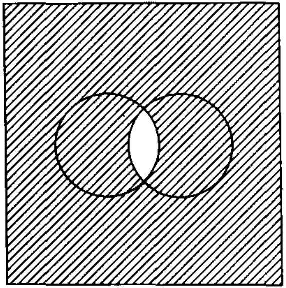


Figura 1: Fonte: Mood et. al (1974)

Teoria de conjuntos: Exemplo



$\overline{A_1 \cup A_2} = \overline{A_1} \cap \overline{A_2}$



$\overline{A_1 \cap A_2} = \overline{A_1} \cup \overline{A_2}$

Figura 2: Fonte: Mood et. al (1974)

Teoria de conjuntos: Exemplo

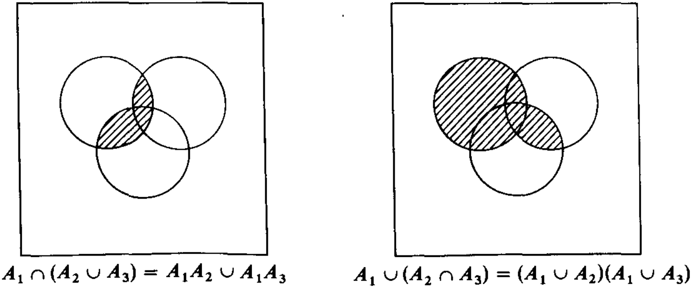


Figura 3: Fonte: Mood et. al (1974)

Teoria de conjuntos: Teoremas

Teorema Leis comutativas $A \cup B = B \cup A$ e $A \cap B = B \cap A$.

Teorema Leis associativas $A \cup (B \cup C) = (A \cup B) \cup C$ e $A \cap (B \cap C) = (A \cap B) \cap C$.

Teorema Leis distributivas $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ e $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.

Teorema $(A^c)^c = A$.

Teoria de conjuntos: Teoremas

Teorema $A \cap \Omega = A$; $A \cup \Omega = \Omega$; $A \cap \phi = \phi$; e $A \cup \phi = A$.

Teorema $A \cap A^c = \phi$; $A \cup A^c = \Omega$; $A \cap A = A$; e $A \cup A = A$.

Teorema Leis de Morgan $(A \cup B)^c = A^c \cap B^c$ e $(A \cap B)^c = A^c \cup B^c$.

Teorema $A - B = A \cap B^c$.

Teoria de conjuntos: Demonstração

Como ilustração de demonstração dos teoremas, mostraremos que $(A \cup B)^c = A^c \cap B^c$, conhecida como Lei de Morgan. Por definição, dois conjuntos são iguais se um está contido no outro.

- *(ida)* Mostremos que $(A \cup B)^c \subset A^c \cap B^c$. Se $\omega \in (A \cup B)^c$ implica $\omega \notin A \cup B$, que implica $\omega \notin A$ e $\omega \notin B$; isto é $\omega \in A^c$ e $\omega \in B^c$, ou seja, $\omega \in A^c \cap B^c$.
- *(volta)* Mostremos que $A^c \cap B^c \subset (A \cup B)^c$. Seja $\omega \in A^c \cap B^c$ que significa que ω pertence a ambos A^c e B^c. Então $\omega \notin A \cup B$, pois, se pertencesse, ω deveria pertencer a pelo menos um dentre A ou B, contradizendo o fato de que ω percente a ambos A^c e B^c. Contudo, $\omega \notin A \cup B$ significa $\omega \in (A \cup B)^c$, completando a prova.

Teoria de conjuntos: Definições

Definição União e intersecção de conjuntos Seja Λ um conjunto de índices e $\{A_\lambda : \lambda \in \Lambda\} = \{A_\lambda\}$ uma coleção de subconjuntos de Ω indexados por Λ.

- O conjunto que consiste de todos os pontos que pertencem a A_λ para pelo menos um λ é chamado de *união* dos conjuntos $\{A_\lambda\}$ e é denotado por $\bigcup\limits_{\lambda \in A_\lambda} A_\lambda$.
- O conjunto que consiste de todos os pontos que pertencem a A_λ para todo λ é chamado de *intersecção* dos conjuntos $\{A_\lambda\}$ e é denotado por $\bigcap\limits_{\lambda \in A_\lambda} A_\lambda$.

Teoria de conjuntos: Teorema

Teorema De Morgan Seja Λ um conjunto de índices e $\{A_\lambda\}$ uma coleção de subconjuntos de Ω indexado por Λ. Então

(i)

$$\left(\bigcup_{\lambda \in A_\lambda} A_\lambda\right)^c = \bigcap_{\lambda \in A_\lambda} A_\lambda^c.$$

(ii)

$$\left(\bigcap_{\lambda \in A_\lambda} A_\lambda\right)^c = \bigcup_{\lambda \in A_\lambda} A_\lambda^c.$$

Teoria de conjuntos: Definições e Teorema

Definição Disjuntos ou mutualmente exclusivos Os subconjuntos A e B de Ω são *mutualmente exclusivos* ou *disjuntos* se $A \cap B = \phi$. Os subconjuntos $A_1, A_2, \ldots$ são *mutualmente exclusivos* se $A_i \cap A_j = \phi$, para todo $i \neq j$.

Teorema Se A e B são subconjuntos de Ω, então (i) $A = (A \cap B) \cup (A \cap B^c)$, e (ii) $(A \cap B) \cap (A \cap B^c) = \phi$.

Prova: (i) $A = A \cap \Omega = A \cap (B \cup B^c) = (A \cap B) \cup (A \cap B^c)$. (ii) $(A \cap B) \cap (A \cap B^c) = (A \cap A) \cap (B \cap B^c) = A\phi = \phi$.

Teorema Se $A \subset B$, então $A \cap B = A$, e $A \cup B = B$.

Definições de espaço amostral e evento

Espaço amostral e evento

Definição O *espaço amostral*, denotado por Ω, é a coleção de todos os resultados possíveis de um experimento.

Definição Um *evento* é um subconjunto de um espaço amostral. A classe de todos os eventos associados com um dado experimento é definido como o *espaço de evento*.

O espaço de eventos é denotado por uma letra latina, geralmente $\mathcal{A}$, $\mathcal{B}$ ou $\mathcal{F}$.

Espaço amostral e evento: exemplo

Exemplo No lançamento de um dado há seis resultados para a face virada para cima.

Assim

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

Seja o evento $A = \{\text{sai número par}\} = \{2, 4, 6\}$.

Para este experimento o espaço amostral é finito e o espaço de eventos compreende todos os subconjuntos de Ω. Há $2^6 = 64$ eventos em $\mathcal{A}$.

Espaço amostral e evento: exemplo

Exemplo O experimento é registrar o número de mortes no trânsito no estado do Paraná no próximo ano.

Qualquer número não negativo é um resultado plausível para este experimento, assim $\Omega = \{0, 1, 2, \ldots\}$.

$A = \{\text{menos de 500 mortes}\} = \{0, 1, 2, \ldots, 499\}$ é um evento.
$A = \{\text{exatamente } i \text{ mortes}\}, i = 0, 1, \ldots$ é um evento elementar.

Há um número infinito de pontos no espaço amostral, e cada ponto é um evento (elementar). Cada subconjunto de Ω é um evento.

Espaço amostral e evento: exemplo

Exemplo Selecione uma lâmpada e registre o tempo em horas que ela fica acesa antes de queimar.

Qualquer número não negativo é um resultado plausível deste experimento, e assim $\Omega = \{x : x \geq 0\}$. Para este espaço amostral nem todos os subconjuntos de Ω são eventos.

Contudo, qualquer subconjunto que possa ser exibido será um evento. Por exemplo, seja

$$\begin{aligned} A =& \{\text{a lâmpada funciona por pelo menos } k \text{ horas mas queima antes de } m \text{ horas}\}. \\ =& \{x : k \leq x \leq m\}, \end{aligned}$$

então A é um evento para qualquer $k \leq x \leq m$.

Álgebra de eventos: definição

Embora não definimos formalmente até aqui quais subconjuntos de Ω constituem nosso espaço de eventos $\mathcal{A}$, destacamos algumas propriedades necessárias de $\mathcal{A}$.

Uma classe de subconjuntos de Ω é uma álgebra de eventos se satisfaz as seguintes propriedades

(i) $\Omega \in \mathcal{A}$.
(ii) Se $A \in \mathcal{A}$, então $A^c \in \mathcal{A}$.
(iii) Se A_1 e $A_2 \in \mathcal{A}$, então $A_1 \cup A_2 \in \mathcal{A}$.

Note que a coleção de todos os subconjuntos de Ω satisfaz as propriedades acima.

Álgebra de eventos: exemplo

Exemplo Considere $\Omega = \{1, 2, 3\}$ e as seguintes coleções de subconjuntos:

$$\begin{aligned}\mathcal{F}_1 =& \{\phi, \Omega, \{1\}, \{2, 3\}\};\\ \mathcal{F}_2 =& \{\phi, \Omega, \{1\}, \{2\}, \{1, 3\}, \{2, 3\}\}.\end{aligned}$$

Vamos verificar que $\mathcal{F}_1$ é uma álgebra:

- $\Omega \in \mathcal{F}_1$, logo a propriedade (i) está atendida.
- Todos os complementares estão em $\mathcal{F}_1$, pois

  $$\phi_c = \Omega; \quad \Omega^c = \phi; \quad \{1\}^c = \{2, 3\} \text{ e } \{2, 3\}^c = \{1\}.$$

  Logo $\mathcal{F}_1$ satisfaz (ii).
- Verificamos que todas as uniões possíveis e seus complementos estão em $\mathcal{F}_1$. Note que $\{1\} \cup \{2, 3\} = \Omega \in \mathcal{F}_1$. Logo, (iii) está atendida e $\mathcal{F}_1$ é uma álgebra.

Álgebra de eventos: teoremas

Teorema

$$\phi \in \mathcal{A}.$$

Prova Pela propriedade (i) $\Omega \in \mathcal{A}$; por (ii) $\Omega^c \in \mathcal{A}$; mas $\Omega^c = \phi$, logo $\phi \in \mathcal{A}$.

Teorema Se A_1 e $A_2 \in \mathcal{A}$, então $A_1 \cap A_2 \in \mathcal{A}$.

Prova A_1^c e $A_2^c \in \mathcal{A}$; logo $A_1^c \cup A_2^c$ e $(A_1^c \cup A_2^c)^c \in \mathcal{A}$, mas $(A_1^c \cup A_2^c)^c = A_1 \cap A_2$ pela Lei de Morgan.

Teorema Se $A_1, A_2, \ldots, A_n \in \mathcal{A}$, então $\bigcup_{i=1}^{n} A_i \in \mathcal{A}$ e $\bigcap_{i=1}^{n} A_i \in \mathcal{A}$.

Prova Segue por indução.

Definição de probabilidade

Função de probabilidade: definição

Definição Uma função de probabilidade $P(\cdot)$ é uma função de conjuntos com domínio na álgebra de eventos $\mathcal{A}$ e contradomínio no intervalo $[0, 1]$ que satisfaz os seguintes axiomas de Kolmogorov:

(i) $P(A) \geq 0$ para todo $A \in \mathcal{A}$.

(ii) $P(\Omega) = 1$.

(iii) Se $A_1, A_2, \ldots$ é uma sequência de eventos mutualmente exclusivos[1] em $\mathcal{A}$ e se $A_i \cup A_2 \cup \ldots = \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$, então

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

[1] isto é, $A_i \cap A_j = \phi$ para $i \neq j; i, j = 1, 2, \ldots$

Função de probabilidade: observação

Muitos autores assumem que o domínio da função de conjuntos é uma sigma-álgebra (σ-álgebra) ao invés de simplesmente álgebra.

Para uma álgebra, temos a propriedade

$$\text{se } A_1 \text{ e } A_2 \in \mathcal{A}, \text{ então } A_1 \cup A_2 \in \mathcal{A}.$$

Numa σ-álgebra, a propriedade acima é substituída por

$$\text{Se } A_1, A_2, \ldots, A_n, \ldots \in \mathcal{A}, \text{ então } \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}.$$

Uma σ-álgebra é uma álgebra, mas não vale necessariamente o oposto.

Função de probabilidade: observação

Se o domínio de uma função de probabilidade é uma σ-álgebra, então o axioma (iii) acima pode ser simplificado por

$$\text{se } A_1, A_2, \ldots \text{ é uma sequência de eventos mutualmente exclusivo em } \mathcal{A},$$
$$\text{então } P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

Um importante teorema, chamado *Teorema da Extensão de Caratheodory*, afirma que se uma função de probabilidade é definida em uma álgebra, então ela pode ser estendida para uma σ-álgebra conveniente.

Importante: $\mathcal{A}$ não pode ser tomada sempre como o conjunto das partes de Ω. A razão é que para Ω suficientemente grande é impossível definir uma função de probabilidade consistente com os axiomas acima.

Propriedades de $P(\cdot)$

Teorema

$$P(\phi) = 0.$$

Prova Tome $A_1 = \phi, A_2 = \phi, A_3 = \phi, \ldots$. Pelo axioma (iii)

$$P(\phi) = P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) = \sum_{i=1}^{\infty} P(\phi),$$

que só é válido se $P(\phi) = 0$.

Propriedades de $P(\cdot)$

Teorema Se $A_1, \ldots, A_n$ são eventos mutualmente exclusivos em $\mathcal{A}$, então

$$P(A_1 \cup \ldots \cup A_n) = \sum_{i=1}^{n} P(A_i).$$

Prova Tome $A_{n+1} = \phi, A_{n+2} = \phi, \ldots$ Então

$$P\left(\bigcup_{i=1}^{n} A_i\right) = P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) = \sum_{i=1}^{n} P(A_i).$$

Propriedades de $P(\cdot)$

Teorema Se A é um evento em $\mathcal{A}$, então

$$P(A^c) = 1 - P(A).$$

Prova $A \cup A^c = \Omega$ e $A \cap A^c = \phi$, logo $P(\Omega) = 1 = P(A \cup A^c) = P(A) + P(A^c)$.

Teorema Se A e $B \in \mathcal{A}$, então $P(A) = P(A \cap B) + P(A \cap B^c)$ e $P(A - B) = P(A \cap B^c) = P(A) - P(A \cap B)$.

Prova $A = (A \cap B) \cup (A \cap B^c)$ e $(A \cap B) \cap (A \cap B^c) = \phi$. Logo $P(A) = P(A \cap B) + P(A \cap B^c)$.

Propriedades de $P(\cdot)$

Teorema Para quaisquer dois eventos A e $B \in \mathcal{A}$,
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$ De forma geral, para eventos $A_1, A_2, \ldots, A_n \in \mathcal{A}$

$$P(A_1 \cup A_2 \cup \ldots \cup A_n) = \sum_{j=1}^{n} P(A_j) - \sum_{i<j}\sum P(A_i \cap A_j) + \sum_{i<j<k}\sum\sum P(A_i \cap A_j \cap A_k) - \ldots + (-1)^{n+1} P(A_1 \cap A_2 \ldots \cap A_n).$$

Prova $A \cup B = A \cup (A^c \cap B)$ e $A \cap (A^c \cap B) = \phi$, logo
$P(A \cup B) = P(A) + P(A^c \cap B) = P(A) + P(B) - P(A \cap B)$. A forma geral segue por indução.

Propriedades de $P(\cdot)$

Teorema Se A e $B \in \mathcal{A}$ e $A \subset B$, então $P(A) \leq P(B)$.

Prova $B = (B \cap A) \cup (B \cap A^c)$ e $B \cap A = A$. Assim, $B = A \cup (B \cap A^c)$, e $A \cap (B \cap A^c) = \phi$; logo $P(B) = P(A) + P(B \cap A^c)$. A conclusão segue pois $P(B \cap A^c) \geq 0$.

Teorema (desigualdade de Boole) Se $A_1, A_2, \ldots, A_n \in \mathcal{A}$, então $P(A_1 \cup A_2 \cup \ldots A_n) \leq P(A_1) + P(A_2) + \ldots + P(A_n)$.

Prova $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) \leq P(A_1) + P(A_2)$. A prova deve ser completada usando indução.

Espaço de probabilidade: definição

Definição Um *espaço de probabilidade* é a trinca $(\Omega, \mathcal{A}, P(\cdot))$, em que Ω é um espaço amostral, $\mathcal{A}$ é uma álgebra de eventos, e $P(\cdot)$ é uma função de probabilidade com domínio $\mathcal{A}$.

Os três componentes estão relacionados. $\mathcal{A}$ é uma coleção de subconjuntos de Ω, e $P(\cdot)$ é é uma função que tem $\mathcal{A}$ como domínio.

Espaço amostral finito com pontos igualmente prováveis: exemplo

Exemplo Considere o exemplo de lançar dois dados (ou um dado duas vezes). Seja $\Omega = \{(i_1, i_2) : i_1 = 1, \ldots, 6; i_2 = 1, \ldots, 6\}$. Aqui i_1 = número na face superior do primeiro dado, e i_2 = número na face superior do segundo dado. Há $6 \times 6 = 36$ pontos amostrais. Atribuiremos probabilidade $1/36$ a cada ponto. (Ω pode ser mostrado como um *lattice* figura a seguir.)

Defina $N(A)$ como o tamanho, ou número de pontos, de A e seja A_7 = evento em que a soma vale 7, então $A_7 = \{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\}$. Logo, $N(A_7) = 6$ e $P(A_7) = N(A_7)/N(\Omega) = 6/36 = 1/6$.

Similarmente $P(A_j)$ pode ser calculado para A_j = total vale j, $j = 1, 2, \ldots, 12$. Como o número de pontos no evento A pode ser facilmente contado, $P(A)$ pode ser calculada para qualquer evento A.

Espaço amostral finito com pontos igualmente prováveis: exemplo

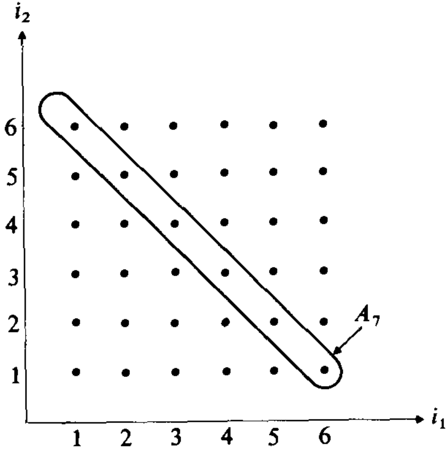


Figura 4: Fonte: Mood et. al (1974)

Espaço amostral finito sem pontos igualmente prováveis: exemplo

Exemplo Considere um experimento com N resultados, digamos $\omega_1, \omega_2, \ldots, \omega_N$, em que sabe-se que o resultado ω_{j+1} tem o dobra da probabilidade do resultado ω_j, $j = 1, \ldots, N-1$, isto é, $p_{j+1} = p_j$, em que $p_i = P(\{\omega_i\})$..

Encontre $P(A_k)$, em que $A_k = \{\omega_1, \omega_2, \ldots, \omega_k\}$. Como

$$\sum_{i=1}^{N} p_j = \sum_{j=1}^{N} 2^{j-1} p_1 = p_1(1 + 2 + 2^2 + \ldots + 2^{N-1}) = p_1(2^N - 1),$$

temos $p_1 = \frac{1}{2^N - 1}$ e $p_j = 2^{j-1}/(2^N - 1)$. Logo,

$$P(A_k) = \sum_{j=1}^{k} p_j = \sum_{j=1}^{k} \frac{2^{j-1}}{2^N - 1} = \frac{2^k - 1}{2^N - 1}.$$

Probabilidade condicional e independência

Probabilidade condicional: definição

Assuma um espaço de probabilidade $(\Omega, \mathcal{A}, P(\cdot))$. Dados dois eventos A e B, queremos definir a probabilidade condicional do evento A dado que o evento B ocorreu.

Definição Sejam A e B dois eventos em $\mathcal{A}$ definidos no mesmo espaço de probabilidade $(\Omega, \mathcal{A}, P(\cdot))$. A *probabilidade condicional* do evento A dado o evento B, denotada por $P(A|B)$, é definida por

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad \text{se } P(B) > 0,$$

e não especificada se $P(B) = 0$.

Probabilidade condicional: exemplo

Exemplo Seja Ω um espaço amostral finito, $\mathcal{A}$ a coleção de todos os subconjuntos de Ω e $P(\cdot)$ a função de probabilidade em que todos os N pontos amostrais $\omega_1, \omega_2, \ldots, \omega_n$ de Ω são igualmente prováveis, ou seja, $P(\{\omega_1\}) = P(\{\omega_2\}) = \ldots = P(\{\omega_N\})$.

Seja $N = N(\Omega)$ o tamanho do conjunto Ω. Para eventos A e B,

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{N(A \cap B)/N}{N(B)/N},$$

em que $N(B)$ é o tamanho do conjunto B.

Para qualquer espaço finito com pontos amostrais igualmente prováveis, os valores de $P(A|B)$ são definidos para quaisquer dois eventos A e B sempre que $P(B) > 0$.

$P(\cdot|B)$ satisfaz os axiomas

Para um dado evento B para o qual $P(B) > 0$, $P(\cdot|B)$ é uma função de probabilidade tendo $\mathcal{A}$ como domínio e satisfaz os três axiomas:

1. $P(A|B) = P(A \cap B)/P(B) \geq 0$ para todo $A \in \mathcal{A}$.
2. $P(\Omega|B) = P(\Omega \cap B)/P(B) = P(B)/P(B) = 1$.
3. Se $A_1, A_2, \ldots$ é uma sequência de eventos mutualmente exclusivos em $\mathcal{A}$ e se $\cup_{i=1}^{\infty} A_i \in \mathcal{A}$, então

$$P\left(\bigcup_{i=1}^{\infty} A_i | B\right) = \frac{P\left(\left(\bigcup_{i=1}^{\infty} A_i\right) \cap B\right)}{P(B)} = \frac{P\left(\bigcup_{i=1}^{\infty} (A_i \cap B)\right)}{P(B)}$$

$$= \frac{\sum_{i=1}^{\infty} P(A_i \cap B)}{P(B)} = \sum_{i=1}^{\infty} P(A_i|B).$$

Propriedades de $P(\cdot|B)$

Teorema

$$P(\phi|B) = 0.$$

Teorema Se $A_1, \ldots, A_n$ são eventos mutualmente exclusivos em $\mathcal{A}$, então

$$P(A_1 \cup \ldots \cup A_n|B) = \sum_{i=1}^{n} P(A_i|B).$$

Teorema Se A é um evento em $\mathcal{A}$, então

$$P(A^c|B) = 1 - P(A|B).$$

Propriedades de $P(\cdot|B)$

Teorema Se A_1 e $A_2 \in \mathcal{A}$, então $P(A_1|B) = P(A_1 \cap A_2|B) + P(A_1 \cap A_2^c|B)$.

Teorema Para quaisquer dois eventos A_1 e $A_2 \in \mathcal{A}$, então $P(A_1 \cup A_2|B) = P(A_1|B) + P(A_2|B) - P(A_1 \cap A_2|B)$.

Teorema Se A_1 e $A_2 \in \mathcal{A}$ e $A_1 \subset A_2$, então $P(A_1|B) \leq P(A_2|B)$.

Teorema Se $A_1, A_2, \ldots, A_n \in \mathcal{A}$, então $P(A_1 \cup A_2 \cup \ldots \cup A_n|B) \leq \sum_{i=1}^{n} P(A_i|B)$.

Teorema de probabilidade total

Teorema Para um dado espaço de probabilidade $(\Omega, \mathcal{A}, P(\cdot))$, se $B_1, B_2, \ldots, B_n$ é uma coleção de eventos mutuamente disjuntos em $\mathcal{A}$ satisfazendo $\Omega = \bigcup_{j=1}^{n} B_j$ e $P(B_j) > 0$ para $j = 1, \ldots, n$, então para todo $A \in \mathcal{A}$,

$$P(A) = \sum_{j=1}^{n} P(A|B_j)P(B_j).$$

Corolário Para um dado espaço de probabilidade $(\Omega, \mathcal{A}, P(\cdot))$, seja $B \in \mathcal{A}$ satisfazendo $0 < P(B) < 1$, então para todo $A \in \mathcal{A}$,

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c).$$

Obs: o teorema acima é válido se $n = \infty$.

Teorema de Bayes

Teorema Para um dado espaço de probabilidade $(\Omega, \mathcal{A}, P(\cdot))$, se $B_1, B_2, \ldots, B_n$ é uma coleção de eventos mutuamente disjuntos em $\mathcal{A}$ satisfazendo $\Omega = \bigcup_{j=1}^{n} B_j$ e $P(B_j) > 0$ para $j = 1, \ldots, n$, então para todo $A \in \mathcal{A}$ para o qual $P(A) > 0$,

$$P(B_k|A) = \frac{P(A|B_k)P(B_k)}{\sum_{j=1}^{n} P(A|B_j)P(B_j)}.$$

Obs: o teorema é válido se $n = \infty$.

Teorema de Bayes

Corolário Para um dado espaço de probabilidade $(\Omega, \mathcal{A}, P(\cdot))$, sejam A e $B \in \mathcal{A}$ satisfazendo $P(A) > 0$ e $0 < P(B) < 1$, então

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)}.$$

Teorema de Bayes: exemplo (DeGroot e Schervich; Magalhães)

Exemplo Considere o problema de avaliar a eficiência de um teste para detectar uma doença. Chamamos de *falso-positivo* ao erro em que o teste indica positivo para um paciente que não tem a doença, e *falso-negativo* se o teste não aponta a doença num paciente doente.

Imagine um teste que resulta positivo para não doentes com probabilidade 0,1. Também com probabilidade 0,1 o teste será negativo para pacientes doentes. Se a incidência da doença na população é de 1 para cada 10 mil habitantes, qual é a probabilidade de uma pessoa estar realmente doente se o teste deu positivo?

Note que, estando ou não doente, existe uma probabilidade não nula do teste indicar a presença da doença.

Teorema de Bayes: exemplo (DeGroot e Schervich; Magalhães)

Exemplo (cont) Sejam os eventos

$$D: \text{a pessoa está doente;}$$
$$A: \text{o teste é positivo.}$$

Temos as informações: $P(D) = 0,0001$; $P(A|D^c) = 0,1$ e $P(A^c|D) = 0,1$. E, assim, $P(D^c) = 0,9999$ e $P(A|D) = 0,9$. Pelo Teorema de Bayes,

$$P(D|A) = \frac{P(A|D)P(D)}{P(A|D)P(D) + P(A|D^c)P(D^c)} = \frac{0,9 \times 0,0001}{0,9 \times 0,0001 + 0,1 \times 0,9999} = 0,0009.$$

Embora pequena, esta probabilidade (cerca de 1 em mil) é cerca de dez vezes a probabilidade da doença na população (1 em 10 mil).

Regra da multiplicação

Teorema Para um dado espaço de probabilidade $(\Omega, \mathcal{A}, P(\cdot))$, sejam $A_1, A_2, \ldots, A_n$ eventos em $\mathcal{A}$ para os quais $P(A_1 \cap A_2 \cap \ldots \cap A_n) > 0$, então

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap \ldots \cap A_{n-1}).$$

Independência de eventos

Definição Para um dado espaço de probabilidade $(\Omega, \mathcal{A}, P(\cdot))$, sejam A e B dois eventos em $\mathcal{A}$. Dizemos que os eventos A e B são *independentes* se, e somente se, qualquer uma das seguintes condições é satisfeita

(i) $P(A \cap B) = P(A)P(B)$.
(ii) $P(A|B) = P(A)$ se $P(B) > 0$.
(iii) $P(B|A) = P(B)$ se $P(A) > 0$.

Independência de eventos: demonstração

Prova Para mostrar a igualdade das três condições, basta mostrar que (i) implica (ii), (ii) implica (iii), e (iii) implica (i).

- Se $P(A \cap B) = P(A)P(B)$, então $P(A|B) = P(A \cap B)/P(B) = P(A)P(B)/P(B) = P(A)$, para $P(B) > 0$; assim (i) implica (ii).
- Se $P(A|B) = P(A)$, então $P(B|A) = P(A|B)P(B)/P(A) = P(A)P(B)/P(A) = P(B)$, para $P(A) > 0$ e $P(B) > 0$; assim (ii) implica (iii).
- Se $P(B|A) = P(B)$, então $P(A \cap B) = P(B|A)P(A) = P(B)P(A)$, para $P(A) > 0$; assim (iii) implica (i).

Note que $P(A \cap B) = P(A)P(B)$, se $P(A) = 0$ ou $P(B) = 0$.

Independência de eventos: exemplo

Exemplo Considere o experimento de lançar dois dados. Sejam os eventos $A = \{\text{o total é impar}\}$, $B = \{\text{sai o número 1 no primeiro lançamento}\}$, e $C = \{\text{o total é sete}\}$.

- A e B são independentes?
- A e C são independentes?
- B e C são independentes?

Temos $P(A|B) = 1/2 = P(A)$, $P(A|C) = 1 \neq P(A) = 1/2$, e $P(C|B) = 1/6 = P(C) = 1/6$. Assim, A e B são indepententes, A não é independente de C, e B e C são independentes.

Independência de eventos: teorema

Teorema Se A e B são dois eventos definidos em um dado espaço de probabilidade $(\Omega, \mathcal{A}, P(\cdot))$, então A e B^c são independentes, A^c e B são independentes, e A^c e B^c são independentes.

Prova $P(A \cap B^c) = P(A) - P(A \cap B) = P(A) - P(A)P(B) = P(A)\left[1 - P(B)\right] = P(A)P(B^c)$

De forma similar para os outros.

Probabilidade à priori e à posteriori: exemplo (DeGroot e Schervich; Magalhães)

Exemplo Suponha que uma máquina produza itens defeituosos com probabilidade p $(0 < p < 1)$ e não defeituosos com probabilidade q igual a $1 - p$. Considere que foram selecionados 6 itens aleatoriamente da produção e os resultados são independentes. Vamos obter, inicialmente, a probabilidade de dois dentre os 6 itens serem defeituosos.

O espaço amostral é composto por todas as sequências dos 6 itens, em que cada item pode ser defeituoso ou não. Represente por D_i e N_i se o item i é defeituoso ou não, respectivamente. Por exemplo, a sequência $(N_1 D_2 N_3 N_4 D_5 N_6)$ indica itens defeituosos na segunda e quinta seleção e não defeituosos nas demais.

Pela independência entre os resultados, temos que

$$P(N_1 D_2 N_3 N_4 D_5 N_6) = p^2 q^4.$$

Probabilidade à priori e à posteriori: exemplo (DeGroot e Schervich; Magalhães)

Exemplo (cont) Essa probabilidade é a mesma para qualquer sequência com dois itens defeituosos, logo:

$$P(\text{dois defeituosos}) = \binom{6}{2} p^2 q^4.$$

Suponha que p é desconhecida mas toma um dos dois valores $p = 0,01$ (operação normal) ou $p = 0,40$ (operação fora do padrão). Sejam B_1 e B_2 os eventos correspondentes aos funcionamentos normal e fora do padrão, respectivamente.

Logo, para qualquer item i,

$$P(D_i|B_1) = 0,01 \quad \text{e} \quad P(D_i|B_2) = 0,40.$$

Probabilidade à priori e à posteriori: exemplo (DeGroot e Schervich; Magalhães)

Exemplo (cont) A independência entre os resultados se torna independência condicional. Em particular, dada a ocorrência de B_1, temos

$$P(N_1 D_2 N_3 N_4 D_5 N_6 | B_1) = P(N_1|B_1)P(D_2|B_1)\cdots P(N_6|B_1).$$

Assim,

$$P(\text{dois defeituosos}|B_1) = \binom{6}{2} 0,01^2 0,99^4 = 1,44 \times 10^{-3};$$

$$P(\text{dois defeituosos}|B_2) = \binom{6}{2} 0,40^2 0,60^4 = 0,311.$$

Probabilidade à priori e à posteriori: exemplo (DeGroot e Schervich; Magalhães)

Exemplo (cont) Considere agora que atribuímos probabilidades às escolhas de p. Isso é o que chamamos de probabilidade *à priori* para p. Essas probabilidades expressam nosso conhecimento quanto ao comportamento da máquina e são baseadas, em geral, na experiência anterior.

Assuma que as probabilidades de funcionamento normal e fora do padrão são, respectivamente, $P(B_1) = 0,90$ e $P(B_2) = 0,10$.

Se na amostra de 6 itens dois foram defeituosos, qual seria a *probabilidade posterior* para o evento B_1? Em outras palavras, qual a probabilidade de a máquina estar funcionando normalmente se observarmos 2 itens defeituosos?

Probabilidade à priori e à posteriori: exemplo (DeGroot e Schervich; Magalhães)

Exemplo (cont) Pelo teorema de Bayes

$$
\begin{aligned}
P(B_1|\text{dois defeituosos}) &= \frac{P(\text{dois defeituosos}|B_1)P(B_1)}{\sum_{i=1}^{2} P(\text{dois defeituosos}|B_i)P(B_i)} \\
&= \frac{1,44 \times 10^{-3} \times 0,9}{1,44 \times 10^{-3} \times 0,9 + 0,311 \times 0,1} = 0,040.
\end{aligned}
$$

Observe o efeito da informação da amostra:

- A probabilidade *à priori* de B_1 era $0,90$.
- A probabiliade *à posteriori* foi consideravelmente reduzida para $0,040$.
- Isso reflete o fato da ocorrência de dois itens defeituosos ser muito mais provável quando B_2 acontece, do que quando ocorre o evento B_1.

Independência de vários eventos: definição

Definição Para um dado espaço de probabilidade $(\Omega, \mathcal{A}, P(\cdot))$, sejam $A_1, A_2, \ldots, A_n$ eventos em $\mathcal{A}$. Dizemos que os eventos $A_1, A_2, \ldots, A_n$ são *independentes* se, e somente se,

$$\begin{aligned} P(A_i \cap A_j) =& P(A_i)P(A_j), \quad \text{para } i \neq j \\ P(A_i \cap A_j \cap A_k) =& P(A_i)P(A_j)P(A_k), \quad \text{para } i \neq j, j \neq k, i \neq k \\ &\vdots \\ P\left(\bigcap_{i=1}^{n} A_i\right) =& \prod_{i=1}^{n} P(A_i). \end{aligned}$$

Independência dois a dois não implica independência

Exemplo Considere o experimento de lançar dois dados. Sejam os eventos $A_1 = \{\text{face ímpar no primeiro dado}\}$, $A_2 = \{\text{face ímpar no segundo dado}\}$, e $A_3 = \{\text{o total é ímpar}\}$.

Temos

- $P(A_1)P(A_2) = 1/2 \times 1/2 = P(A_1 \cap A_2)$,
- $P(A_1)P(A_3) = 1/2 \times 1/2 = P(A_3|A_1)P(A_1) = P(A_1 \cap A_3)$, e
- $P(A_2)P(A_3) = 1/4 = P(A_2 \cap A_3)$.

Assim, A_1, A_2 e A_3 são independentes dois a dois. Contudo, $P(A_1 \cap A_2 \cap A_3) = 0 \neq 1/8 = P(A_1)P(A_2)P(A_3)$. Logo, A_1, A_2 e A_3 não são independentes.